DOCUMENT RESUME

ED 332 755 JC 910 274

AUTHOR Rubadeau, Duane O.; Rubadeau, Ronald J.
TITLE Guidelines for Grading.
INSTITUTION College of New Caledonia, Prince George (British Columbia). Centre for Improved Teaching.
PUB DATE May 83
NOTE 37p.
PUB TYPE Guides - Classroom Use - Teaching Guides (For Teacher) (052)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; Community Colleges; *Evaluation Criteria; Foreign Countries; *Grades (Scholastic); *Grading; Higher Education; Scores; *Scoring; Statistical Analysis; *Student Evaluation; *Testing Problems; Two Year Colleges

ABSTRACT

Designed to assist community college instructors in the difficult task of grading students, this seven-part guidebook provides philosophical, theoretical and practical advice. Section I covers common problems associated with grading, the need for grades, and the link between grades and subsequent performance. Section II, "Limitations of Grades," addresses the tendency for grades to lack clear definition and to be unreliable; and examines absolute versus relative grading systems. In section III, "Grading Achievement Versus Related Factors," grades based on improvement and criterion referenced grading are explored. Section IV, "The Single Versus the Multiple Grading System," looks at the number of grades on the grade scale, whether letters or numbers should be used to denote grades, and the meaning of grades. Section V, "Basing Grades on Composite Scores," is followed by sections providing a short bibliography and a glossary of terms. (JMC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED332755

RUBADEAU & RUBADEAU

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

K. Plett

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☑ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Guidelines For Grading

BEST COPY AVAILABLE

JC910274

College of New Caledonia. Centre for Improved Teaching
Prince George, British Columbia
May 1983

PREFACE

Probably the most difficult task that instructors face is assigning grades to students or making evaluation statements about their performance. There are many critics of the current grading practices used in the educational setting, even though schools and colleges must serve an evaluation role. The question is not - Should we evaluate? - but rather, How should evaluation statements be derived and communicated?

Most serious instructors are well aware of the limitations and dangers of grading. Here, we are referring to the problems such as: the tremendous variation in assigning grades, the lack of a clear definition of what the grading system means, and the lack of objective data upon which grades are assigned. The purpose of the Guideline is not to attempt to resolve these problems, but rather to help instructors arrive at a system of grading and reporting grades that the instructors can feel comfortable using in the light of the demands that are placed on the instructor.

The Guideline is comprised of seven sections. Section I deals with the Introductory Material related to the problems involved in Grading. Section II covers the limitations of Grades and Grading Systems. Grading Achievement versus Related Factors is discussed in Section III. Section IV deals with the Single Versus the Multiple Grading System. Section V looks at the procedure for Basing Grades on Composite Scores. The bibliographing is contained in Section VI and the Glossary in Section VII.

We thank Linda Fieguth for her excellent work on typing the manuscript and Diane Jacobs for her always beautiful work on the cover design.

Duane O. Rubadeau

Ronald J. Rubadeau

May 1983.

TABLE OF CONTENTS

SECTION		PAGE
	Preface	1
I	Introduction	1
	Some Problems Involved In Grading	2
	The Need For Grades	3
	Grades And Subsequent Performance	6
II	Limitations Of Grades	8
	Lack Of Clearly Defined Grades	9
	Grades Tend To Be Unreliable	10
	Absolute Versus Relative Grading Systems	11
III	Grading Achievement Versus Related Factors	13
	Grades Based On Improvement	14
	Criterion Referenced Grading	16
IV	The Single Versus The Multiple Grading System	17
	The Number Of Grades On The Grade Scale	19
	Should Letters Or Numbers Be Used To Denote Grades?	20
	The Meaning Of Grades	20
V	Basing Grades On Composite Scores	22
VI	Bibliography	29
VII	Glossary	31

SECTION I

INTRODUCTION

SOME PROBLEMS INVOLVED IN GRADING

The problem of grading student achievement has been a difficult one at all levels of education. On a yearly basis there is a phenomenal coverage by the media as well as professional journals of articles that either criticize current educational practices or point out some new method to replace existing methods.

The first problem involved in grading is that measurements which involve human behaviour are subject to errors that are due primarily to three factors: the imperfections of the units on the measuring scales; the complexity of the measurements to be made; and the lack of consistency in the facts to be measured. Grades are measurements of educational achievement, hence, they are subject to the three varieties of errors.

A second problem in grading is that grading systems tend to become main issues in the area of educational controversy. For example, the orientation of the uniqueness of the student and the student's need for reassurance, led to criticisms of the orientations requiring competitive pressures and common standards of achievement for all students. On the other hand, the emphasis on basic education and pursuing academic excellence has raised the hue and cry for more formal evaluations of achievement and more vigorous standards of attainment. My goodness, there is that ugly word - standards. That is a word that every Canadian should have tatooed on the inside of his or her eyelids. Why? Because there are many things in our national life which are in direct opposition to standards - mediocrity, complacency, desire for making a fast dollar, reluctance to criticize poor work, and our fondness for short-cuts. Every Canadian knows we have to come to terms with these shortcomings sooner or later.

The third area in which grading systems present continuing problems is that

they require instructors to stand in judgment over their students. This is not seen as a role in a friendly two-way interaction and may well result in anti-social feelings. It is easy to give students a good grade, especially if it is higher then the grades they expected. However, there will probably be more instances of disappointment than there will be of pleasure in grading.

It is not likely that a system of grading will be found that will make the process easy and painless. We are not saying that present grading systems are beyond improvement, but rather, that new grading systems, however they are devised and followed are not likely to solve the basic problems of grading. The need is not for new grading systems, as they are available now. The problem seems to be that the more confident instructors feel they are doing a good job of grading, the less likely they are to be aware of the difficulties of grading, such as: the personal biases that may be reflected in their grades, and the fallibility of their judgements.

The Need For Grades

At all levels of education, most instructors go along with the idea that grades are necessary. Once in awhile a scream of protest is heard, pointing out that grading is a vicious practice and should be eliminated, however, there is no way to demonstrate that abolishing grades will produce better achievement. The only way you can determine whether achievement is better under one set of conditions rather than another is to measure it. When you eliminate the measurement component you have no way of comparing the two approaches. No matter what level of education you are involved with, the comparison of achievement between students appears to be inevitable. Children in Grade 5 will ask each other how they fared on the spelling test. College and university students do the same thing. What appears to be the issue in grading is not the use of grades, but rather, the misuse of grades.

Grades have a very wide variety of uses: First, grades are used as self-evaluation measures to let the student know where he/she stands. Second, they are used in making educational decisions and career planning. Third, grades are used to indicate the student's performance to other educational or training agencies, as well as to potential employers.

As you are well aware, education is a very expensive operation. As a result, we need to monitor each student's performance as accurately and carefully as possible, in order to attain the maximum performance from our students and to get the best possible use from the facilities we have available. Hence, grades serve the function of letting us know whether our students are learning and to what extent they are learning.

Grades also have the function of reinforcing, stimulating and directing the student learning. This happens to be one of the controversial areas related to the use of grades. There are a number of people at the various education levels that feel that grades provide reinforcements that are artificial and the motivation of the student is under the control of other people, namely the instructors. This is true, however, so are a lot of the other tangible rewards of achievement. Most instructors experience internal satisfaction from doing a good job and knowing that their students are learning. However, most of us are quite delighted that we do not have to live on these internal or intrinsic reinforcements alone - it is also nice and very rewarding to receive a paycheck and a bonus for work well done. The idea is that we as instructors can not live solely on intrinsic reinforcements, so why should we expect students to do it.

In order for grades to be effective for reinforcing, stimulating and directing student behaviour, the grades have to be valid. That is, the high grades have to go to those students having attained the greatest number of

course objectives. However, the grades we hand out must be based on a broad sampling of the student's performance. Grades that are based on irrelevant or incidental learnings are not only detrimental to the students, but are invalid measures of the attainment of the course objectives.

From time to time, you will hear instructors and students playing down the role of grades with the general orientation that what a student learns is much more important than the grade they receive. This idea appears to be based on the assumption that the relationship between what is learned and the grade received is very low or non-existent. There is another common comment heard in the academic setting that indicates that grades are not an end in themselves, therefore, why should tests, quizzes or examinations be given if they are just used for assigning grades.

Generally, the grade received by a student is not of itself an important educational outcome, but neither are the diplomas or certificates toward which the student is striving. They are however, valid indicators of the educational achievements made by the student to that point. Therefore, the need is to make the goal of best possible educational achievement match the goal of highest possible grades. When the goals of achievement and goals of grades do not match, the problem appears to lie with the instructors teaching the courses and assigning the grades. Grades are necessary and if they are invalid, the solution is not in de-emphasizing grades, but rather, in assigning grades with greater care so they are representative of the degree of achievement attained by the student. Hence, we feel that instructors should take greater pains to improve the validity of the grades they assign, instead of wasting their time looking for a painless substitute method of grading.

GRADES AND SUBSEQUENT PERFORMANCE

Many of the critics of grading systems refer to the studies indicating low relationships between grades attained and subsequent performance. It is little wonder that current grading practices lead to conclusions such as: high grades do not always predict future performance accurately, or low grades do not invariably indicate the student will fail in future endeavours. There are several reasons for the low relationships reported between grades and subsequent performance. One reason is that while learning (as measured by the grades assigned) is a condition for future performance, other factors such as motivation, opportunity and just plain old dumb luck have a great deal to do with future performance. A second reason for the low relationship between grades and subsequent performance appears to be the lack of accurate measuring instruments for assessing achievement. This occurs when the instructor doesn't have the ability or is not willing to take the time to do an accurate job of measuring and reporting achievement. The third and final reason for the low relationship between grades and subsequent performance is the very difficult problem of defining an acceptable level of success for subsequent performances by the student.

Training and education are expected to make a positive contribution to the student's future performance. Unless something is really drastically wrong, instructional programs are developed and designed to aid students in learning what they have to know in order to succeed in subsequent situations. Grades then, should indicate the extent to which students have learned the objectives we have set for our courses. As a result, if the grades have a low relationship to subsequent success, there has to be a problem either in the assessment of what students have learned or a problem with the program of instruction or both. Therefore, for grades to be poorly related to subsequent

performance is not rational or tolerable and hence, not acceptable.

SECTION II

LIMITATIONS OF GRADES

The major limitations of grades as they are distributed by many colleges are that there is no clearly defined, generally accepted definition of what the grades mean and a lack of objective data to use as a foundation for assigning grades.

LACK OF CLEARLY DEFINED GRADES

This limitation is centred around the fact that the meaning of grades and grading standards vary greatly from course to course and from one instructor to another. Further, the problem is compounded by instructor biases which helps to reduce the validity of the grades.

Numerous articles in the Journal of Educational Measurement and Journal of Educational Psychology have pointed out the fantastic variability in grading standards and practices running from elementary school right on through to the graduate school setting. For example, when the common five letter system (A,B,C,P & F) is used, the percent of students receiving A grades ran from 0 - 40%, for those students receiving a B grade from 10 - 50%, and for those students receiving an F grade from 0 - 25%.

As a way of trying to get instructors organized in their approach to grading, some colleges publish a summary of the grades assigned in various courses and by the different instructors. About all that is accomplished is a great deal of screaming as to what instructor(s) appear to be the easy touch for a grade. What is usually omitted from these published summaries of grades is whether the instructor was using a well organized set of objectives, whether the instructor opted out and graded on the curve, or whether the course had a very high applied content which had to be transferred and utilized from one learning component to another.

The lack of a clearly defined basis for grading standards and meaning of grades makes it easy for biases to enter into the grading policy and thus lower

the validity of the grades still further. Here we deal with such nebulous factors such as appearance, sociability and skill in verbal expression. Writing or oral presentation ability of the student should not influence the grade for a particular course if they have not met the objectives. Often, however, the student who writes well and has a good line will get good grades even though they do not have a clue as to the subject matter for the course. Data gathered over the years indicates that women students are more likely to get higher grades than men students of the same ability and achievement. Also, students who are liked by the instructor tend to get higher grades than students of the same ability and achievement level who were not well liked. We have also run across instructors who use high grades as rewards and low grades as punishments for behaviours completely outside of the realm of attaining educational objectives. The net result of this state of affairs is that students tend to have a great deal of evidence to support their contention that particular instructors are extremely unfair in their grading policy.

GRADES TEND TO BE UNRELIABLE

Back in 1912-13, Starch and Elliott published a series of studies on the unreliability of teacher's grades in the areas of English, Geometry and History. All of the English teachers were given an identical copy of an English examination paper and told to grade them on the basis of 100% for perfection. The grades assigned to the paper ranged from 50 to 98%. They found similar results for the grading of history and geometry papers. Similar results are found by students who get a P or F from one instructor, have a friend turn in the identical paper to another instructor and it receives a C grade. What this means, is that the grading of single examination papers is not very reliable.

ABSOLUTE VERSUS RELATIVE GRADING SYSTEMS

In general, two kinds of grading systems have evolved here in Canada. In the early 1900's nearly all grading was in terms of percent. Hence, a student who learned everything that was demanded of him/her would receive a grade of 100%. The cutoff score for a minimally acceptable performance was usually set around the 70% level. As the grade was based on the student's learning of the material and his/her performance did not depend on any other student's grade, the system was referred to as an absolute grading system.

The second kind of grading system that evolved is based on the letter grades. Usually the five letters (A,B,C,P & F) are employed. In this system, the A indicates outstanding achievement, B is for above average achievement, C is for average achievement, P indicates below average achievement and F indicates the person has not achieved sufficiently to obtain credit for the course. In this system, the letter grade indicates a student's achievement in relation to the achievement of his/her fellow students. As a result of this comparison in performance, this system is referred to as a relative grading system.

There are variations to the relative grading system, with the most popular being referred to as grading on the "curve". The curve is the graphic portrayal of the normal distribution. One procedure for grading on the curve is to estimate the percent of grades that should fall into the five categories of your grading system. These estimates are based on the theoretical normal curve. With this approach, the highest 10 percent of the scores get a grade of A and the lowest 10 percent get a grade of F. The next highest 20 percent get a grade of B and the next lowest 20 percent get a grade of P. The middle 40 percent of the scores get a grade of C. For some instructors, the preceding

variations of the relative grading system is too cut and dried and totally lacking in imagination. Hence, another variation of the relative grading system evolved, which appears to have greater credibility from the application of statistics. In this procedure the instructor sets the upper and lower limits for each grade level by applying the mean and standard deviation to the test scores. As an example, those students with a score 1.5 standard deviations or more obove the mean receive a grade of A. Those students with a score 1.5 standard deviations or less below the mean recieve a grade of F. Those students with a score between .5 and 1.5 standard deviations above the mean get a grade of B, while those with a score between .5 and 1.5 standard deviations below the mean recieve a grade of P. Those students with scores in the middle of the distribution, that is between .5 standard deviations above the mean and .5 standard deviations below the mean recieve a grade of C. While each of these approaches have a certain percent of the students receiving A's and F's, the second approach is a bit more flexible in that it does not have a set percentage receiving those grades.

At the present time, most instructors tend to use letter grades, however, many arrive at the letter grade by converting the grades from percentages over to the letter grade system.

SECTION III

GRADING ACHIEVEMENT VERSUS RELATED FACTORS

As mentioned previously, some instructors base the grades they give on various aspects of student behaviour that are not directly related to the attainment of the instructional objectives. This is especially true when the instructor does not have a set of instructional objectives that are given out to the students. It is quite likely that these factors not directly related to achievement will continue to be utilized, especially when they have been found to be useful in controlling student behaviour.

The prime requirement of a good grading system is that the grades must give the most accurate indication of the extent to which the student has attained the objectives in the course. If the improvement of student motivation or attitudes is one of the instructional objectives, then it is reasonable that changes in motivation or attitude be taken into account when assigning grades. When attitudinal or motivational changes are not a direct part of the instructional objectives, they should be omitted from the process of determining the student's grade.

Grades Based on Improvement

As a way of enhancing the accuracy or fairness of their grades, some instructors have based their grades on the improvement the student has exhibited, rather than comparing a student's performance to the performance of the rest of the students in the course. This particular approach involves the assessment of entry level skills and abilities, usually with some type of pre-test. The differences between these scores and the scores on a post-test (final examination) are used to indicate the degree of improvement for each of the students. The major problem is that these measures of improvement are often not reliable. What is needed to obtain a reliable and valid measure of improvement is the development of two forms (parallel forms) of the same test. That is, you have to develop two tests that measure the same content, with the

same degree of difficulty, using different test items. While not an impossible task, you will certainly enhance your test construction skills. The general idea is that if your pre- and post-tests are reliable and valid, the differences in student achievement of instructional objectives may be used as an indicator of the effectiveness of instruction.

While the measurement of improvement appears to provide a better method of measuring achievement, there are some other problems beyond the reliability of the test scores. For example, in some cases, knowing the students' status in relation to the rest of the class can be of greater value than knowing how much he/she has learned from the course. That is, how did the student learn in relation to the other students? Faster? Slower? About average? Further, you will always have the situation where the students who received a low grade on the pre-test having the greatest probability of showing huge gains in achievement than the students who had relatively high scores on the pre-test. As students, contrary to the veiws of many of their instructors, do not live in a vacuum, they very soon realize that under this system of improvement, the idea is to start out with a very low score, so the gains will be large when measured over the length of the course.

In spite of the disadvantages of grading on the basis of improvement, one real advantage of this grading system is that it gives all of the students a better opportunity to earn good grades. In the comparison approach to grading, the generalization is that students attaining high grades in one course tend to get high marks in other courses as well. The converse also holds true, in that students receiving low grades in one course tend to get low grades in the other courses. This in turn leads to feelings of discouragement and reduced motivation, which in turn, produces still poorer performance.

Another factor that we have overlooked, especially at the college level, is

to make sure that the students that enter particular courses or programs have the requisite skills and abilities needed to attain the instructional objectives for that course or program. Of course, once this task is taken into account and implemented, it will certainly have a devastating impact on reducing the number of P and F grades an instructor is able to give out. In fact, when all of the students have the requisite skills and abilities to successfully attain the instructional objectives, yet a large percentage of them fail, perhaps it is time to take a look at the instructor and instruction, rather than the students.

Criterion Referenced Grading

Hopefully you are aware that there are phenomenal individual differences in the amount of material that students will learn in nearly every course. These differences may be reduced to some extent by the organization of the course content, but in general, unless you make the course so simple-minded that everyone can grasp everything immediately, you will have to admit there are vast individual differences in learning ability between students.

One of the approaches that has merit in working with students of differing ability and motivational level is the criterion-referenced method. This method does not eliminate individual differences, but it does allow the opportunity for all students to attain the criteria established for assessing achievement. Criterion-referenced programs are designed to have students attain mastery of subject matter at one level before moving ahead to the next level of material. While these programs centered around mastery learning do offer greater opportunities for all students to learn, they certainly do not get rid of the individual differences in ability and motivation.

SECTION IV

THE SINGLE VERSUS THE MULTIPLE GRADING SYSTEM

The grade a student receives for his/her work in any course is, and should be a composite of many factors. The student's achievement is based in part, on the information presented, the understanding of that information as well as the student's interest and motivation to learn. In addition, there are a multiplicity of factors that determine how well and to what extent the student has attained the instructional objectives. Such things as examination scores, completion of assignments, participation, attendance and motivation are all involved in determining the student's grade. With all of these factors involved, how can a letter or a percentage cover all of these aspects of learning? An answer that is becoming more and more common is that a single letter or percentage cannot give an accurate reflection of all the factors involved in the learning.

The net result is that two different orientations to grading have evolved over the past several years. One approach attempts to expand the areas of student development that are being graded. The other approach is centred around an increase in the specific factors of what is being graded. While these two approaches have some value, they are also fraught with some problems. First, these grading systems make grading more difficult, rather than simplifying the task. Second, these grading systems produce problems in terms of coming up with precise definitions of exactly what is being graded. Third and finally, there is the difficulty of securing enough data related to each component of learning to come up with a reliable grade.

The general idea is that multiple grading is not a cure-all for the problems involved in grading, in that this approach may place demands on the instructor, which may be beyond his/her capabilities. That is, the instructor may not be able to gather all the information needed for a multiple grading system. Fortunately there are other possibilities, so we may be able to

improve the grading process without resorting to multiple grading.

The Number of Grades On The Grade Scale

The two grading systems that are commonly employed are percent grades and letter grades. The letter grades came into existence when a number of instructors and teachers realized that the accuracy of the percent grading system was not good enough to warrant the supposed precision of the percent system. About the best that most instructors can do is to distinguish five levels of achievement, hence the shift to the letter grading system.

While the percent and letter grading systems are the most popular, two other systems have evolved over time. These systems collapse grades into two categories. One system is the PASS-FAIL approach. However, the two category pass-fail system appeared to be a bit too restrictive, as plus and minus signs were often added to expand the categories and differentiate between levels of performance. The other grading system that many colleges use for certain classes is an S for satisfactory and a U for unsatisfactory performance.

The idea that grading difficulties can be made less complex and errors reduced by cutting down on the number of categories in the system has a great deal of appeal. The major problem with a two category system is the loss of information for both the student and the instructor. So you have a more precise, easier grading system that provides less information. Thus, by cutting down on the number of categories for grading we reduce errors and increase the precision of the grades we assign. However, the errors now become extremely important. For example, the difference between a C and a P grade is much more critical then the difference between 76 and 78 percent.

Should Letters Or Numbers Be Used To Denote Grades?

When the switch was made from percent grades to letter grades, the letters served to magnify the difference between the relative and the absolute percent grades. Unfortunately, the letter grades have two inherent problems: The first is that the letter grades tend to give the impression that you have made evaluations of achievement, rather than measuring achievement. The second problem with letter grades is that they have to be converted onto some type of numerical scale in order to average them. For example, to compute grade point averages the A is usually equivalent to 4.0 points, the B to 3.0 points, the C to 2.0 points, and so forth.

For these reasons it would appear to be worthwhile to go back to the number system to report grades. Unfortunately, some amount of confusion is encountered with the establishment of sets of letters or numbers that have new or different meaning.

The Meaning Of Grades

The meaning of a grade is determined in two ways: first, by how it is defined; and secondly, by how it is used. For example, if the instructor gives very few P and F grades, not too many C grades, a large number of B grades, and a fair amount of A grades, the average for this instructor is no longer the C grade as established by the college, but rather the average would probably end up in the B- to B range.

There are a number of reasons why instructors tend to deviate from college grading policies: For some instructors, grading is viewed as the personal domain of the instructor, which allows the instructor a tremendous amount of freedom and leeway in assigning marks. For other instructors, the tendency may be to give very few high or low grades, which preserves the C grade as the average, but frustrates the students that are really working hard and end up

with the same grades as those who have done little or nothing in the way of achievement; Finally, some instructors that have provided a clear-cut set of objectives and are grading on the basis of mastery of the subject matter will tend to have a grading distribution that has more high grades than low grades assuming the students are motivated.

SECTION V

BASING GRADES ON COMPOSITE SCORES

When you determine a course grade, you usually do this by combining grades on class participation, papers, and scores on tests and quizzes. Each of these grades carries a different amount of weight for determining the final grade for the course. To obtain grades with the best possible validity, you have to give each grade the proper weight. Your task then, is to determine what those weights are versus what they should be. If there is a great difference between these two sets of grades, the next step is to rectify the disparity.

There are several principles that will be helpful in determining how much each grade influences the final grades for a course:

1. Using several different kinds of measures of competence is better than the use of only a single measure. This assumes that each measure is relevant to the objectives of the course and the behaviors can be measured or observed reliably. For example, excessive use of tests may give an unfair advantage to the students having special test-writing skills and may present a severe handicap to students that show their achievement in discussions, projects, or oral presentations. However, in no way should the ability to be a smooth-talker, personal charm, or self-confidence be mistaken for a good understanding of the material. You also have to be very careful of the amount of weight placed on subjective judgements that cannot be measured reliably.
2. When the measures of achievement are closely related, the problem of assigning weights is much less of a problem than when these measures are not related. For most courses, the measures of achievement are related closely enough so that accurate weighting is not a serious problem. That is, the natural (unweighted) grades in this case, would

provide grades that are nearly as valid as those produced by using a sophisticated statistical procedure.

3. The actual weight that a component of the final grade will carry depends on the variability of its measures and the relationship of these measures with the measures of other components of achievement. This, of course, makes the precise influence of any one measure of achievement very difficult to determine. To gain an approximation of the weight of a measure of achievement, the standard deviation of the measures of the component will serve very nicely. (See Rubadeau - Guide to Elementary Statistics, 2nd ed, Section IV-D.) For example, if one set of grades has a standard deviation twice the size of the standard deviation for another set of grades, the set with the low standard deviation will carry twice the weight of the other set of grades.

The table below shows that the weight of one measure of achievement (scores on exam 1) on a composite (the sum of scores on the three exams) depends on the variability (standard deviation) of the exam scores. The upper portion of the table shows the scores of three students, A, B, and C, on three exams, together with their total scores on the three exams. Student B has the highest total and Student A the lowest total. Moving down the table, the next section indicates the students rank on the three exams. It is interesting to note that each student made the highest score on one exam, the middle score on another, and the lowest score on the third.

Moving down the table still further, the third section provides information about the maximum possible (total points), the mean scores, and the standard deviations of the scores on the three exams. Exam 1 has the highest number of total points. Exam 2 has the highest mean score and Exam 3 has scores with the

largest variability.

WEIGHTED EXAM SCORES

Tests	Exam 1	Exam 2	Exam 3	Total
Student Grades				
A	53	65	18	136
B	50	59	42	151
C	47	71	30	148
Student Ranks				
A	1	2	3	3
B	2	3	1	1
C	3	1	2	2
Exam Characteristics				
Total Points	100	75	50	225
Mean Score	50	65	30	145
Standard Deviation	2.5	5	10	6.5
Weighted Scores	X4	X2	X1	
A	212	130	18	360
B	200	118	42	360
C	188	142	30	360

Now, on which exam was it most important to do well? On which exam was the penalty for ranking last the hardest on the student? The answer is clearly on Exam 3, the exam with the greatest variability of scores. Which test ranked the students in the same order as their final ranking based on total scores? Again the answer is Exam 3. Thus the influence of one aspect on a composite depends not on total points or mean score, but on score variability.

The next task is to figure out how we can get these exam scores to carry

equal weights. This can be accomplished by weighting the scores to make the standard deviations equal. This can be seen in the last section of the table. Scores on Exam 1 are multiplied by 4, to change their standard deviation from 2.5 to 10, the same as the standard deviation on Exam 3. Scores on Exam 2 are multiplied by 2, to change their standard deviation to 10. With equal standard deviations the tests carry equal weight, and give students having the same average rank on the tests the same total scores.

When the whole range of possible scores is used, the score variability is closely related to the extent of range of available scores. In effect, this means that scores on a 50 item objective test are likely to have five times the weight of scores on a 10 point essay test question, assuming that the scores extend across the entire range in both cases. However, if only a small portion of the possible range of scores is used, the length of the exam can be a very poor guide to the variability of scores.

If you, as the instructor, are having some difficulty dealing with what you feel the component weightings ought to be and what they actually are, you have two alternatives.

The first, is to multiply what you feel is the underweighted scores by some weighting factor to increase the variability of these scores and thus increase the weight they carry. The other approach is to increase the number of observations of the underweighted scores, or increase the precision of the measures of the underweighted component, which in turn increases the weight it carries. Although the first method is likely to be more convenient, the second method is likely to yield more reliable and valid grades.

As an example, you have promised students in one of your courses that the final grade for the course will be based on five components and they will have the following weights:

Class Participation	15%
Term Paper	15%
Weekly Quizzes	20%
Midterm Exam	20%
Final Exam	30%

Your task then is to obtain enough independant measures in the area of Class Participation in order that the variability of these scores is about half of the variability of the scores on the final exam. Further, the final exam should be at least 1 1/2 times the length of the midterm exam. That is, if the midterm contains 50 items, the final exam should contain 75 items.

You will do well to warn your students that the actual weight of each grade in a composite grade may differ somewhat from what the intended weight might be. However, if you follow your weighting plan, you can assure your students with some degree of confidence that the deviations that do occur will not have a significant effect on the validity of the grades.

A mistake that is often made by instructors is to convert test grades to letter grades and record the letter grades in their grade book, then reconvert the letter grades to numbers for the purpose of calculating the final average. A much better procedure to follow is to record the exam grades along with other numerical measures directly into the grade book. These grades can be added with their appropriate weights, to obtain a composite grade that can be converted into the student's course grade.

The recording of exam scores, rather than letter grades usually saves time and contributes to accuracy as well. Whenever a distribution of scores is converted to letter grades, some information is lost. Generally this information cannot be retrieved when the letter grades are changed back to numbers. Each C grade, whether a high C or a low C is given the same value in

a reconversion of the grades from letters to numbers. Thus, to avoid the loss of information it is usually desireable to record the raw number grades and not record the grades after conversion to letters.

SECTION VI

BIBLIOGRAPHY

Bibliography

Anastasi, A., Psychological Testing, 4th ed., New York: Macmillan, 1976.

Brown, F., Measuring Classroom Achievement, New York: Holt, Rinehart and Winston, 1981.

Hopkins, C. & Antes, R., Classroom Testing: Administration, Scoring and Score Interpretation, Itaska, Il: Peacock, 1979.

Mehrens, W. & Lehmann, I., Standardized Test In Education, 3rd ed., New York: Holt, Rinehart and Winston, 1980.

Rubadeau, D., Guide to Elementary Statistics, 2nd ed., Prince George, B.C.: Canadian Psychological Services, 1980.

Rubadeau, D., Garrett, W. & Rubadeau, R., Appropriate Testing, Prince George, B.C.: Centre For Improved Teaching, College of New Caledonia, 1983.

SECTION VII

GLOSSARY

GLOSSARY

Achievement Test: a test designed to measure the extent to which a person has acquired certain information or mastered certain skills, usually as the result of specific instruction although this may not always be the case.

Essay Item: a test item requiring the test taker to write a narrative answer in response to a question or problem situation.

Evaluation: judgment of value, quality, or worth of some performance or program.

Grade: the symbol or mark used to evaluate a student's level of performance in a course or on a particular measure, for example A, B, 80%, Pass or Satisfactory.

Item: a single question or exercise on a test.

Learning: a relatively permanent change in performance as a result of motivation, practice and experience.

Mean: the arithmetic average of a set of test scores.

Measurement: the process of assigning numbers to performance according to specified rules and procedures.

Multiple-Choice Item: an incomplete sentence or question followed by several possible choices; the test taker selects the alternative that best completes the statement or answers the question.

Normal Distribution (Curve): the symmetrical bell-shaped distribution with most scores near the center and fewer at the ends.

Objective Scoring: scoring that ensures a high degree of agreement between competent (trained) scorers.

Passing Score: the minimum score a test taker can attain and still pass a test.

Performance Test: a test requiring some physical or psychomotor activity, for instance, playing a saxaphone, typing, or doing modern dance recital.

Pretest: a test given at the beginning of instruction to determine whether students have mastered the prerequisite material, and/or to assess their entry level skills.

Raw Score: the score derived directly from the scoring of the test, for example, number correct, total points, time to complete the task.

Reliability: how consistently a test measures over time, occasions, or samples of items; the degree to which test scores are affected by measurement errors. Measured by a reliability coefficient and the standard error of measurement.

Score: the quantitative value assigned to an individual's performance on a test, subtest, scale, or group of items.

Standard Deviation: a measure of the variability of a set of scores around the mean. The lower the standard deviation, the more the scores cluster around the mean; the higher the standard deviation, the more variable the scores.

Subtest: a set of items administered and scored as a separate portion of a more comprehensive test.

Test: any systematic procedure for measuring a sample of behaviour.

Validity: the degree to which a test measures what it is designed to measure, or predicts some external criterion; major subcategories include content validity, construct validity, and criterion-related validity.

Variability: how widely the scores in a distribution are dispersed around the mean; usually measured by the standard deviation.

ERIC Clearinghouse for Junior Colleges JUL 5 1991